



Revista Iberoamericana de Argumentación

ἐπεὶ δὲ ταύτην τὴν ἐπιστήμην ζητοῦμεν

Director
Luis Vega

Secretaria
Paula Olmos

Edición Digital
Roberto Feltrero

Innovación, Big Data y Epidemiología

David Casacuberta

*Departamento de Filosofía
Universitat Autònoma de Barcelona
E08193 Bellaterra, Barcelona
david.casacuberta@uab.cat*

RESUMEN

El objetivo de este texto es mostrar como la epidemiología ha avanzado gracias a innovaciones en el proceso argumentativo y en cómo la tecnología digital y más concretamente, el fenómeno del *big data* -masivas bases de datos alimentadas en tiempo real con millones de entradas que pueden procesarse para buscar correlaciones de forma rápida- puede ser una poderosa herramienta para seguir innovando en epidemiología, permitiendo trabajar con grandes poblaciones en tiempo real, aumentando así las posibilidades de establecer correlaciones estadísticas lo suficientemente argumentadas para poder inferir la realidad del fenómeno y buscar posibles mecanismos causales con al menos cierta probabilidad de llegar a un resultado fiable.

PALABRAS CLAVE: innovación en procesos argumentativos; epidemiología, e-ciencia.

ABSTRACT

The main aim of this paper is to show how epidemiology has advanced thanks to innovations in the argumentative process and its connections with digital technology. More specifically, the phenomenon of big (data-fed massive databases in real time with millions of entries that can be processed to look for correlations quickly) can be a powerful tool for further innovation in epidemiology, allowing researchers to work with large populations in real time, thus increasing the chances of establishing statistics sufficiently well argued to infer the reality of the phenomenon, and for possible causal mechanisms correlations with at least some probability of reaching a reliable result.

KEYWORDS: innovation in argumentative processes; epidemiology; e-science.



Copyright©DAVID CASACUBERTA

Se permite el uso, copia y distribución de este artículo si se hace de manera literal y completa (incluidas las referencias a la Revista Iberoamericana de Argumentación), sin fines comerciales y se respeta al autor adjuntando esta nota. El texto completo de esta licencia está disponible en: <http://creativecommons.org/licenses/by-nc-sa/2.5/es/legalcode.es>

1. OBJETIVO DE ESTE TEXTO

La epidemiología es una ciencia en un importante proceso de readaptación, tanto sobre su objeto final de estudio como su metodología de trabajo. Parte de esta readaptación está asociada a la dificultad a la hora de establecer nuevos resultados significativos. Tal y como exponía Gary Taubes en su artículo para *Science* "Epidemiology faces its limits" (Taubes 1995) cada vez resulta más complejo establecer conexiones significativas entre variables en la epidemiología de factores de riesgo. Las conexiones son tenues, los resultados no son replicables, no resulta accesible establecer mecanismos causales, etc.

El objetivo de este texto es mostrar como la epidemiología ha avanzado gracias a innovaciones en el proceso argumentativo y en cómo la tecnología digital y más concretamente, el fenómeno del *big data* -masivas bases de datos alimentadas en tiempo real con millones de entradas que pueden procesarse para buscar correlaciones de forma rápida- puede ser una poderosa herramienta para seguir innovando en epidemiología, permitiendo trabajar con grandes poblaciones en tiempo real y aumentando así las posibilidades de establecer correlaciones estadísticas lo suficientemente argumentadas para poder inferir la realidad del fenómeno y buscar posibles mecanismos causales con, al menos, cierta probabilidad de llegar a un resultado fiable.

2. UNA PRIMERA APROXIMACIÓN A LA EPIDEMIOLOGÍA

Aunque tendemos a asociar la epidemiología a enfermedades infecciosas, en realidad la epidemiología tiene un objetivo más amplio: entender las correlaciones entre salud y hábitos de vida (en un sentido amplio). Así, la epidemiología puede estudiar la relación entre prácticas sexuales no seguras y la transmisión del SIDA (volvemos a las enfermedades infecciosas), pero un estudio que conecte deglutir comida basura de forma habitual con un mayor riesgo de enfermedades coronarias también es epidemiología.

A partir del análisis estadístico, comparando un grupo de prueba con uno de control, el epidemiólogo intenta descubrir factores de riesgo para nuestra salud y genera recomendaciones, que luego los políticos pueden aplicar (o no) cuando diseñan políticas sanitarias.

Factor de riesgo se entiende de forma amplia, como cualquier cosa o proceso que pueda poner en peligro nuestra salud. No tiene que ser necesariamente un objeto físico; también puede ser un patrón cultural. Así, es tarea del epidemiólogo contrastar cuál es el peligro real de utilizar móviles de manera sistemática, e incluso puede analizar el efecto que los modelos ultradelgadas tienen como "modelo a seguir" en la distribución de la anorexia.

A medida que la epidemiología avanza, se observa que los factores de riesgo pueden ser cada vez más elaborados y abstractos. Así, la epidemiología social estudia cómo pertenecer a un grupo social determinado puede ser un factor de riesgo para nuestra salud. No se trata simplemente de pertenecer a un grupo con prácticas que incluyen factores de riesgo infecciosos (como drogodependientes que comparten jeringuillas) sino de algo mucho más complejo, sutil y filosóficamente fascinante: e.g. ¿hay correlaciones claras positivas entre sueldo y salud? Y ello genera interesantes quebraderos de cabeza para políticos y epidemiólogos. ¿Podría ser que luchar contra la pobreza no sea solamente una cuestión de política social sino de salud pública?

Una parte significativa de los factores de riesgo que estudia la epidemiología tienen también connotaciones sociales negativas: ser drogodependiente, adicto al sexo, miembro de una minoría étnica o religiosa, etc. Los registros que tengan los epidemiólogos de los grupos en estudio han de estar por tanto celosamente protegidos para evitar filtraciones de datos privados que podrían acabar creando problemas a las personas recogidas en esos registros.

No se trata de una preocupación meramente académica. A mediados de los noventa, el estado de Massachusetts decidió liberar datos "anonimizados" de todas las visitas hospitalarias de los trabajadores públicos. El entonces gobernador William Weld aseguró en una rueda de prensa que la privacidad de esos datos estaba asegurada. La entonces estudiante Latanya Sweeney consiguió una copia de esos datos, la cruzó con el censo electoral y fue capaz, entre otras cosas, de localizar los registros de visitas hospitalarias del propio gobernador Weld (Anderson 2009). Más adelante, Sweeney fue capaz de demostrar (Malin et al 2003) que es posible asociar hasta un 87% de los ciudadanos americanos con sus registros de hospital aunque se eliminen del registro nombre, dirección, etc. con solo la fecha de nacimiento y el código postal, cruzando los datos con el censo electoral.

El *big data* resulta muy significativo para analizar y situar estas cuestiones. Pensemos en *Google Flu Trends* (Cook et al 2011), y en cómo los expertos en *big data* de Google fueron capaces de predecir con bastante detalle la evolución de la

gripe en Estados Unidos a partir de las búsquedas online que los usuarios hacían de términos como "gripe", "síntomas de gripe", "remedios contra la gripe", etc. y asociando esas búsquedas a localizaciones específicas a partir del número de IP del usuario que llevaba a cabo la búsqueda. Los resultados son extraordinarios y las curvas de distribución estadística de la gripe y de búsqueda de la gripe se superponen la mayoría de las veces.

Está claro que datos así podrían ser muy útiles en investigación médica. Desde establecer la evolución de una posible pandemia a averiguar cómo podría ayudar a los epidemiólogos en su investigación poder establecer correlaciones entre comportamiento online y comportamiento físico. También habría que tener en cuenta los peligros que un procesamiento detallado de datos supuestamente neutros podría generar en nuestra privacidad.

En este artículo quiero investigar otra cuestión relacionada: el papel que el *big data* podría tener a la hora de innovar en los procesos argumentativos de la epidemiología y así conseguir establecer y contrastar nuevas hipótesis en campos en los que, en principio, el avance estaría en entredicho. Para entenderlo presentaré en la siguiente sección el concepto de límites epistemológicos a la epidemiología.

3. ¿EL FIN DE LA EPIDEMIOLOGÍA DE FACTORES DE RIESGO?

En su muy citado artículo de *Science* -Taubes (1995)- el experto en nutrición Gary Taubes argumentaba que la época dorada de la epidemiología había terminado. No hay ninguna posibilidad de que aparezca otro John Snow.

Recordemos brevemente los orígenes de la epidemiología. Es un proceso basado en buena parte en innovación argumentativa (Johnson 2007). En el siglo XIX la creencia común sobre la transferencia del cólera era que se trataba de un tipo de "miasmas" que se desplazaban a través del aire. El conocimiento de los micro-organismos era limitado, y no estaba bien establecida su función en la transmisión de enfermedades.

El médico John Snow creía que en realidad el cólera se transmitía a través del agua, pero al no disponer de ningún mecanismo causal capaz de mostrar cómo se llevaría a cabo esa transmisión (el patógeno concreto fue identificado muy posteriormente) procedió a demostrar su teoría de una forma indirecta. Empezó a trazar puntos en un mapa indicando lugares con enfermos de cólera durante la epidemia de Londres en 1854, y pudo mostrar cómo los casos de cólera tendían a

agruparse alrededor de fuentes de agua contaminadas, siendo así capaz de fundamentar la hipótesis de que el cólera se transmite en realidad a través del agua contaminada y no por aire.

Observemos aquí cómo Snow creó la epidemiología no al establecer una teoría sobre los micro-organismos, ni tampoco estaba presentando nuevos datos desconocidos para el resto de los científicos, sino innovando en la forma de presentar los datos: innovando en la forma en que se argumentaba. En paralelo arrancaba una nueva disciplina: la visualización de datos.

No hay ninguna diferencia lógica entre un listado de muertos y enfermos de cólera con sus direcciones, y ponerlos como puntos en un mapa. Pero al intelecto humano le resulta mucho más fácil convencerse de la conexión entre agua contaminada y cólera viéndolos ubicados en un mapa.

Esta ha sido la forma estándar de trabajar en la epidemiología de factores de riesgo. A través de análisis estadísticos los médicos se han convencido, por ejemplo, de que el tabaquismo provocaba cáncer de pulmón, aunque faltara identificar durante mucho tiempo el mecanismo causal específico.

Pues bien, para Taubes, esos días dorados de la epidemiología se estaban acercando a su fin. A partir de ahora, declara Taubes en su artículo, la epidemiología solo puede esperar establecer tenues lazos entre los estilos de vida y las enfermedades -o la falta de éstas. Y los resultados siempre estarán abiertos a las réplicas fáciles y rápidas, con la sospecha de que el factor causal realmente en juego pueda ser algo completamente diferente. No dispondremos de resultados fiables y certeros.

La razón de tal imposibilidad de lanzar nuevas hipótesis en epidemiología no es ni mucho menos que ya tengamos totalmente controlados los factores de riesgo y no sean capaces de generar nuevas enfermedades. Bien lejos de ello. La evidencia circunstancial de que hay factores de riesgo asociados a nuestros hábitos de vida está clara. En las palabras del propio Taubes:

What drives the epidemiologic quest for risk factors is the strong circumstantial evidence that what we eat, drink, breathe, and so on are major factors in many devastating illnesses. Rates of heart disease, for example, have changed much faster over recent decades than can be explained by genetic changes, implicating dietary and environmental causes. And the fact that no single cancer affects every population at the same rate suggests that factors external to the human body causes 70% to 90% of all cancers. (Taubes 1995. p. 164).

¿Cómo establecemos esos factores de riesgo? La manera más razonable para

practicar la epidemiología de factor de riesgo es a partir de la creación de grupos de prueba y control aleatorios. Sin embargo, esos estudios son difíciles de poner en práctica debido a que:

- i) podrían pasar décadas hasta que un factor de riesgo en realidad generara una enfermedad;
- ii) es posible que el investigador tenga que trabajar con miles de personas para poder tener en cuenta todos los factores pertinentes;
- iii) no es ético dar a alguien una sustancia o fomentarle un estilo de vida que creemos que pueda ser perjudicial.

En su lugar, tenemos que contentarnos con comparar el grupo test -las personas con la enfermedad que queremos estudiar, y un grupo control -los que no la tienen- y comprobar qué diferencias existen en el estilo de vida, dieta o factores medio-ambientales entre el control y el grupo de prueba, que puedan ser causalmente relevantes.

Una vez que aceptamos esta limitación, Taubes establece su tesis principal: no hay más conexiones causales relevantes que la epidemiología de los factores de riesgo pueda descubrir, ya que son demasiado complejas e interactúan entre sí en una forma no lineal.

Observemos que la tesis de Taubes no es empírica, sino epistemológica, filosófica. Taubes está poniendo límites epistemológicos al proceso de argumentación en epidemiología. Taubes no niega que no existan mecanismos causales que conecten obesidad y menor esperanza de vida. Nos indica que nunca habrá argumentos suficientemente convincentes para mostrar tal conexión al gusto de todos.

Estos límites de la epidemiología se basan en las siguientes consideraciones:

- i) Ya no es posible identificar el factor causal realmente relevante de varios factores de confusión. Normalmente los hábitos o los factores ambientales van asociados, en un paquete, y es difícil establecer cuál es realmente el causante de la enfermedad. El ejemplo clásico es la conexión entre tomar café y cáncer de pulmón que parece que sería un factor de confusión debido a que la gente que fuma normalmente también toma café. O el epidemiólogo que creyó que trabajar en un tipo de fábrica determinado causaba cáncer y no observó que sus resultados eran causados por el hecho de que todo el mundo que trabajaba en aquella fábrica era más joven que la población media, de modo que era normal que los cánceres aparecieran en gente más joven. (Friedman 2003)

ii) Hay una gran cantidad de sesgos en la forma en que los grupos control y los de prueba se seleccionan. Taubes señala que una forma común de seleccionar a las personas, la marcación en un teléfono de dígitos al azar, no crea grupos muy representativos, ya que por lo general dejan de lado a personas con menos recursos económicos (o bien no contestan el teléfono, ya que tienen dos trabajos y casi nunca están en casa, o están demasiado ocupados para tener tiempo para participar en un estudio).

iii) Aparte de hábitos que resultan fáciles de contar como fumar -los fumadores tienden a saber cuantos cigarrillos, o paquetes, fuman cada día- es difícil para los sujetos experimentales recordar cuántas grasas no saturadas tomaron la semana pasada o lo mucho que caminan.

iv) Las revistas científicas no están interesadas en resultados negativos, por lo que los epidemiólogos tienen que torcer mucho sus números para conseguir algo que suene relevante y por lo tanto publicable. Consideremos por ejemplo el recentísimo debate en la editorial en la revista *Nature* en relación a un artículo epidemiológico que ponía en duda la relación entre sobrepeso y esperanza de vida (Hughes 2013).

En el artículo original, Flegal et al (2013), se ofrecían datos estadísticos que apuntaban a que estar obeso de hecho podía aumentar nuestra esperanza de vida en un 6%. Su publicación indignó a muchos epidemiólogos. Buena parte de las críticas al ya citado estudio de Flegal et al (2013), como Willett et al (2013) consideran que los autores han retocado los datos hasta conseguir ese 6% extra de esperanza de vida para los obesos. Veinte años después y seguimos teniendo el mismo tipo de debate que Taubes pronosticaba. La sombra de Taubes es sin duda alargada, y no hemos podido sacarnos de encima la duda que ponía en 1995 sobre el futuro de la epidemiología.

Desde una perspectiva de racionalización de recursos es triste pensar cuántos equipos han repetido un mismo estudio que no ofrece resultados positivos porque equipos anteriores no han conseguido publicarlo.

v) La prensa tiende a presentar resultados muy preliminares como verdad indiscutible y ello daña la credibilidad de la investigación epidemiológica.

En la sección anterior mencionamos la epistemología social. Si la epistemología clásica de factor de riesgo presentaría según Taubes grandes problemas para producir nuevos resultados, ¿qué podría pasar si queremos vincular categorías abstractas como clase social o las prácticas religiosas a la salud y la enfermedad?

Consideremos por ejemplo estas correlaciones. Estadísticamente, en Estados Unidos el votante republicano tiende a considerarse más feliz que el demócrata. ¿Se trata de sus creencias religiosas? Si es así:

- i) ¿cuáles de estas creencias son realmente relevantes?
- ii) ¿Ser religioso es causalmente relevante para la salud?

O, en la misma línea: la práctica de la meditación budista está asociada en muchos estudios epidemiológicos con el bienestar y buena salud. Potenciaría por lo visto el sistema inmunológico. Pero esta afirmación es más difícil de testar de lo que parece. Por ejemplo: ¿cómo sabemos si alguien es de hecho budista? La práctica de la meditación está abierta a muchas interpretaciones del término "budista". ¿Qué elementos de la práctica budista son realmente relevantes?

También tenemos el problema del grupo de prueba: ¿Las personas que meditan son más felices o las personas felices tienen más capacidad para meditar? No todo el mundo puede permitirse el lujo de dedicar un tiempo cada día a meditar. La meditación sería así una práctica de clase media/acomodada con lo que aparecerían aquí factores de confusión.

Para demostrar que Taubes se equivocaba y que la epidemiología de los factores de riesgo puede seguir adelante, tenemos que seguir innovando en la forma en que se recoge y procesa la evidencia. Aquí el big data podría ser de gran ayuda para lograrlo.

4. ¿QUÉ ENTENDEMOS POR *BIG DATA*?

Por primera vez en la historia somos capaces de generar bases de datos con decenas de millones de entradas, ya sea a partir de datos de sensores para obtener información sobre procesos naturales o por la acción combinada de cientos de miles de usuarios en procesos de creación colectiva en Internet (Casacuberta 2003). Al mismo tiempo, disponemos de una impresionante capacidad informática para procesar esos datos prácticamente en tiempo real y poder extraer todo tipo de conclusiones.

(Manyika et al. 2011)

Estas nuevas posibilidades cambian algunas de las reglas del juego, muchas de ellas activas desde el Renacimiento y algunas de ellas desde el Paleolítico. La filosofía actual debería asumir, analizar y reflexionar sobre este nuevo paisaje epistemológico, ético, sociológico y cultural al que nos lleva inevitablemente el *big data*, esta nueva e impresionante capacidad de recabar y procesar millones de datos simultáneamente. Especialmente en relación a estas tres áreas:

Privacidad. Hasta la aparición del *big data*, la dispersión de los datos que dejábamos sobre nosotros mismos en el mundo era tan alta que podíamos estar confiados en que nadie podría reconstruir nuestros gustos, intenciones, deseos, temores, acciones legales e ilegales a no ser que fuera un psicópata lunático que nos siguiera a todas horas. En la sección anterior ya hemos hablado del problema que puede representar el análisis estadístico de datos médicos. Pero las posibilidades van mucho más allá. Eche un vistazo a la historia de búsquedas que Google guarda de usted y pregúntese por un momento que podría hacer un *cracker* malintencionado, un gobierno dictatorial o una aseguradora sin escrúpulos con ella. De forma alegre dejamos fragmentos de nuestra identidad en Facebook, Twitter, Amazon, Flickr, etc. y no somos conscientes de lo fácil que es reconstruir nuestras vidas a partir de todos esos rastros digitales que hemos ido dejando (Boiller y Firestone 2010).

Tendencias globales. Hemos hablado ya de la impresión que genera ver lo bien correlacionados que van la evolución de la gripe y las búsquedas sobre "gripe" en Google. Ese día, Brin y Page nos trajeron una impresionante herramienta para poder estudiar tendencias y actitudes sociales de una manera que hasta ahora nunca nos habríamos imaginado. De hacer estimaciones a partir de unos pocos cientos de llamadas telefónicas a disponer de centenares de millones de acciones de usuarios que no son resultado de una encuesta explícita, sino expresión implícita de sus verdaderos intereses y actitudes a partir de aquello que buscan en la red. No es de extrañar que Barack Obama se haya traído al Ala Oeste a un equipo especializado en *data mining* para descubrir qué preocupa a los votantes norteamericanos y cómo responder mejor a sus quejas y preocupaciones, sus anhelos y esperanzas (Baker 2009). Esta tendencia es altamente relevante para la epidemiología, y podemos pensar en cómo una disciplina como la epidemiología podría mejorar ostensiblemente al poder acceder a datos de búsqueda para establecer mejor el movimiento de una enfermedad, o cómo se correlaciona con determinados hábitos, pertenencia a clase social, etc.

e-pistemología. Los tiempos en que Einstein podía revolucionar la física sentado en un prado escribiendo ecuaciones en un cuaderno han pasado a la historia. Pensemos por un momento en los millones de datos que genera en milésimas de segundo un acelerador de partículas. Esos datos ya no puede procesarlos una persona, ni siquiera un equipo, y se hace necesario procesar esos datos con un ordenador. Por no hablar de las demostraciones matemáticas asistidas por ordenador -como la del teorema de cuatro colores, que ya no puede revisarse a mano de forma manejable o de los supercomputadores que simulan *in silico* los segundos posteriores al Big Bang. Las fronteras entre ciencias formales y ciencias experimentales se difuminan y el ordenador pasa de ser una mera ayuda a ser una pieza imprescindible e insustituible de la investigación científica.

5. CONCLUSIÓN

Las implicaciones a veces son directas. Otras más sutiles, pero no está de más prepararnos para un mundo bastante diferente al que estamos acostumbrados. Para bien y para mal, las tres tendencias descritas arriba son altamente relevantes para el desarrollo de la epidemiología. En un futuro bien cercano vamos a ver la aparición de la e-epidemiología, una disciplina capaz de saltar las objeciones planteadas en Taubes (1995) para desarrollar nuevas y complejas correlaciones basadas en trabajos donde el proceso de analizar los datos y extraer conclusiones, aunque esté mediado finalmente por un humano, reposa en el procesamiento directo por parte de grandes supercomputadores que funcionan como una caja negra de la que simplemente hemos de confiar en el resultado (Vallverdú 2009).

Esta última implicación es, pienso, la filosóficamente más relevante. Apunta a cómo la innovación de las tecnologías digitales puede llegar a transformar la forma en que entendemos una argumentación, entendiéndola más como un proceso mediado entre personas y máquinas, siguiendo la llamada "segunda ola" de la mente extendida (Sutton 2006).

En síntesis, esta nueva interpretación apuntaría a una lectura de los procesos argumentativos que no tienen lugar meramente en la cabeza, sino que combinan pensamiento, lenguaje, texto escrito y los resultados generados por una computadora digital. El papel que tales dispositivos digitales pueden jugar en una demostración en epidemiología es compleja, y puede llevarnos a imaginar un cambio en las reglas de juego de la metodología científica.

Sin embargo, si aceptamos la idea de que los procesos cognitivos y argumentativos no tienen por qué tener lugar exclusivamente en la cabeza, no necesitamos abdicar del método científico tradicional. Cuando apareció la demostración del teorema de los cuatro colores -el teorema de que para pintar cualquier mapa de manera que dos regiones adyacentes no tengan el mismo color, con cuatro colores es suficiente- los filósofos se alarmaron, pensando que la matemática se iba a convertir en una ciencia experimental, *a posteriori* (Kripke 1980, Tymoczko 1979).

Es posible releer estas derivaciones de la matemática -o la epidemiología- como resultado de una mente extendida (Rufener 2011; Casacuberta y Vallverdú, en prensa). El mecanismo es simple: no es más que la extensión y sistematización del uso de diagramas que son comunes en las demostraciones matemáticas y que actúan como elementos externos a la mente que ayudan a construir la prueba. Los datos generados por el gigante de Google pueden resultar mucho más sorprendentes y difíciles de calcular, pero el mecanismo es, finalmente, el mismo.

REFERENCIAS

- Anderson, N. (2009) "Anonymized" data really isn't—and here's why not" en *Ars Technica* (<http://arstechnica.com/tech-policy/2009/09/your-secrets-live-online-in-databases-of-ruin/>) [última revisión 28-07-2013]
- Baker, S. (2009) "What Data Crunchers Did for Obama". *BusinessWeek*, Enero 23, 34-38.
- Bollier, D., & Firestone, C. M. (2010). *The promise and peril of big data*. Washington, DC, USA: Aspen Institute, Communications and Society Program.
- Casacuberta, D. (2003) *Creación colectiva*. Barcelona: Gedisa.
- Casacuberta, D. y Vallverdú, J. (En prensa) "E-Science and the Data Deluge". *Philosophical Psychology*.
- Cook S, Conrad C, Fowlkes A. L., Mohebbi M.H. (2011) "Assessing Google Flu Trends Performance in the United States during the 2009 Influenza Virus A (H1N1) Pandemic." *PLoS ONE* 6(8).
- Flegal, K. M. ; Kit, B. K. ; Orpana, H. ; Graubard, B. I. "Association of All-Cause Mortality With Overweight and Obesity Using Standard Body Mass Index Categories. A Systematic Review and Meta-analysis" *Journal of the American Medical Association*. Vol 309 N.1 71-82.
- Friedman, G. (2003) *Epidemiology: A primer*. NYC: McGraw Hill Medical.
- Hughes, V. "The Big Fat Truth", *Nature* Vol 497 Num 7450. pp. 2-3.
- Johnson, S. (2007) *The Ghost Map: The Story of London's Most Terrifying Epidemic--and How It Changed Science, Cities, and the Modern World*. NYC: Riverhead Trade.
- Kripke, S. (1980). *Naming and Necessity*. Cambridge: Harvard University Press.
- Malin, B., Sweeney, L. y Newton, E. (2003) "Trail Re-Identification: Learning Who You Are From Where You Have Been" en *LIDAP-WP12*. Carnegie Mellon University, Laboratory for International Data Privacy, Pittsburgh, PA: March 2003. pp. 48-56
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., y Byers, A. H. (2011). *Big data: The next frontier for innovation, competition, and productivity*. Atlanta: McKinsey Global Institute.

- Rufener, C.M. (2011). "The four-color theorem solved, again: extending the extended mind to the philosophy of mathematics." *Res Cogitans*, 2, 215-221.
- Sutton, J. (2006) "Exograms and interdisciplinarity: History, the extended mind and the civilizing process". En Menary (ed) *The extended mind*. Farham: Ashgate.
- Vallverdú, J. (2009). "Computational epistemology and e-science. A new way of thinking". *Minds and Machines*, 19, 557-567.
- Taubes, G. (1995) "Epidemiology faces its limits". *Science* vol 269 14 de julio 1995 pp. 164-169
- Tymoczko, T. (1979). The four-color problem and its philosophical significance. *The Journal of Philosophy*, 76, 57-83.
- Willet, W., Hu, F. B., Thun, M. (2013) "Overweight, Obesity and All-cause Mortality" *The Journal of the American Medical Association* Vol 309 N. 16 1679

AGRADECIMIENTOS: Esta investigación ha sido financiada por el gobierno español desde el proyecto de la DGICYT: FFI2011-23238, "Innovación en las prácticas científicas: acercamientos cognitivos y sus consecuencias filosóficas".

DAVID CASACUBERTA: Es profesor agregado en el Departamento de Filosofía de la Universidad Autónoma de Barcelona. Su línea de investigación actual es el efecto de las tecnologías digitales en los procesos de innovación en ciencia.